# RPM Support - Issue #1548

## published errata contain packages not in repo

01/18/2016 10:44 PM - mhrivnak

| | | | |
|---|---|---|---|
| **Status:** | CLOSED - CURRENTRELEASE | **Start date:** | |
| **Priority:** | High | **Due date:** | |
| **Assignee:** | semyers | **Estimated time:** | 0:00 hour |
| **Category:** | | | |
| **Sprint/Milestone:** | | | |
| **Severity:** | 2. Medium | **Groomed:** | No |
| **Version:** | Master | **Sprint Candidate:** | No |
| **Platform Release:** | 2.8.1 | **Tags:** | Pulp 2 |
| **OS:** | | **Sprint:** | |
| **Triaged:** | Yes | **Quarter:** | |

**Description**

When an erratum has packages for multiple repos, for example el6 and el7, the sync operation will concatenate those packages onto the same erratum unit. At publish time, it appears that the whole package list is being published, which confuses yum on the client. For example an el6 client will see both el6 and el7 packages as available updates.

From the BZ:

```
Steps to Reproduce:
1. Synchronize multiple channels (RHEL5, RHEL6 and RHEL7)
2. Register a RHEL6 system to *only* the RHEL6 base channel
3. `yum updateinfo cve all`

Actual results:
When the errata info is displayed it lists rpms from the RHEL5 and RHEL7 channels.  I.e:

[root@rhel6host ~]# yum updateinfo cve all
Loaded plugins: boks-protect, changelog, fastestmirror, security, tmprepo, verify, versionlock
org-rhel-6                                | 2.1 kB     00:00
org-rhel-6/primary                        |  993 B     00:00
org-rhel-6/updateinfo                     | 1.9 kB     00:00
i  CVE-2014-7169 Important/Sec. bash-3.2-33.el5_11.4.x86_64
i  CVE-2014-7186 Important/Sec. bash-3.2-33.el5_11.4.x86_64
i  CVE-2014-7187 Important/Sec. bash-3.2-33.el5_11.4.x86_64
i  CVE-2014-7169 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
i  CVE-2014-7186 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
i  CVE-2014-7187 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
   CVE-2014-7169 Important/Sec. bash-4.2.45-5.el7_0.4.x86_64
   CVE-2014-7186 Important/Sec. bash-4.2.45-5.el7_0.4.x86_64
   CVE-2014-7187 Important/Sec. bash-4.2.45-5.el7_0.4.x86_64
updateinfo list done

Expected results:
Should only display info relevant to the system.  i.e:

[root@rhel6host ~]# yum updateinfo cve all
Loaded plugins: boks-protect, changelog, fastestmirror, security, tmprepo, verify, versionlock
org-rhel-6                                | 2.1 kB     00:00
org-rhel-6/primary                        |  993 B     00:00
org-rhel-6/updateinfo                     | 1.9 kB     00:00
i  CVE-2014-7169 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
i  CVE-2014-7186 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
i  CVE-2014-7187 Important/Sec. bash-4.1.2-15.el6_5.2.x86_64
updateinfo list done
```

**Related issues:**

| | |
|---|---|
| Is duplicate of RPM Support - Issue #1366: Advisory package list doesn't matc... | **CLOSED - CURRENTRELEASE** |

## Associated revisions

**Revision 73d67dd4 - 03/15/2016 12:08 AM - semyers**

published updateinfo only contains units in repo

Errata units in Pulp contain all units in all repos that are linked to errata with the same id, which was resulting in published errata referencing packages that weren't actually available in the published repo. This limits packages in published errata updateinfo XML to only the packages that are contained in the published repo.

fixes #1366 https://pulp.plan.io/issues/1366

fixes #1548 https://pulp.plan.io/issues/1548

## History

**#1 - 01/19/2016 09:47 PM - mhrivnak**

Probably we need to filter packages out of the package list at publish time.

**#2 - 01/29/2016 05:44 PM - jortel@redhat.com**

*- Platform Release changed from 2.8.0 to 2.8.1*

**#3 - 03/01/2016 12:05 AM - semyers**

*- Status changed from NEW to ASSIGNED*

*- Assignee set to semyers*

**#4 - 03/04/2016 12:56 AM - semyers**

In a pulp_rpm PR[0], errata were converted to append packages to existing errata from different repos, rather than overwrite the existing errata unit. That was a needed improvement, and related PRs can also be seen in the bugzilla for that[1]. On the publish side, though, this means that we're now listing every unit pulp knows about in a repo's errata, when published, instead of only the units associated with that repo. It is simple enough to apply this filter when publishing a repo, but since errata package lists consist of package names (and thus only their NEVRA), in order to associate a package from an errata with a given pulp repo, we must first determine that unit that corresponds with that package, and then find see if a repo/unit association exists for that unit in the repo being published. If so, include it in updateinfo, if not, don't.

mhrivnak wrote:

> Probably we need to filter packages out of the package list at publish time.

Here's a diff where I give that a shot using some minimally-invasive surgery:

```
diff --git a/plugins/pulp_rpm/plugins/distributors/yum/metadata/updateinfo.py b/plugins/pulp_rpm/plugins/distr
ibutors/yum/metadata/updateinfo.py
index 532c9fa..fdce5f6 100644
--- a/plugins/pulp_rpm/plugins/distributors/yum/metadata/updateinfo.py
+++ b/plugins/pulp_rpm/plugins/distributors/yum/metadata/updateinfo.py
@@ -2,6 +2,7 @@ import os
 from xml.etree import ElementTree

 from pulp.plugins.util.metadata_writer import XmlFileContext
+from pulp.server.db.model.criteria import UnitAssociationCriteria

 from pulp_rpm.plugins.distributors.yum.metadata.metadata import REPO_DATA_DIR_NAME
 from pulp_rpm.yum_plugin import util
@@ -13,9 +14,10 @@ UPDATE_INFO_XML_FILE_NAME = 'updateinfo.xml.gz'


 class UpdateinfoXMLFileContext(XmlFileContext):
-    def __init__(self, working_dir, checksum_type=None):
+    def __init__(self, working_dir, checksum_type=None, conduit=None):
         metadata_file_path = os.path.join(working_dir, REPO_DATA_DIR_NAME,
                                           UPDATE_INFO_XML_FILE_NAME)
+        self.conduit = conduit
         super(UpdateinfoXMLFileContext, self).__init__(
             metadata_file_path, 'updates', checksum_type=checksum_type)

@@ -95,13 +97,22 @@ class UpdateinfoXMLFileContext(XmlFileContext):
             name_element.text = pkglist['name']

             for package in pkglist['packages']:
-
```

```
                package_attributes = {'name': package['name'],
                                      'version': package['version'],
                                      'release': package['release'],
                                      'epoch': package['epoch'] or '0',
                                      'arch': package['arch'],
                                      'src': package.get('src', '') or ''}
+
+                # If an rpm with this unit key isn't associated with this repo,
+                # don't include in this repository's updateinfo XML
+                package_key = package_attributes.copy()
+                del(package_key['src'])
+                criteria = UnitAssociationCriteria(type_ids=['rpm'], unit_filters=package_key)
+
+                if self.conduit is not None and not self.conduit.get_units(criteria):
+                    continue
+
                 package_element = ElementTree.SubElement(collection_element, 'package',
                                                          package_attributes)
```

```
diff --git a/plugins/pulp_rpm/plugins/distributors/yum/publish.py b/plugins/pulp_rpm/plugins/distributors/yum/
publish.py
index 2df7cbb..a9e7ada 100644
--- a/plugins/pulp_rpm/plugins/distributors/yum/publish.py
+++ b/plugins/pulp_rpm/plugins/distributors/yum/publish.py
@@ -562,7 +562,8 @@ class PublishErrataStep(platform_steps.UnitModelPluginStep):
         one that is built into the UpdateinfoXMLFileContext
         """
         checksum_type = self.parent.get_checksum_type()
-        self.context = UpdateinfoXMLFileContext(self.get_working_dir(), checksum_type)
+        self.context = UpdateinfoXMLFileContext(self.get_working_dir(), checksum_type,
+                                                self.get_conduit())
         self.context.initialize()
         # set the self.process_unit method to the corresponding method on the
         # UpdateInfoXMLFileContext as there is no other processing to be done for each unit.
```

...and it totally works! The only problem is that it makes publishing errata go **extremely** slowly when I test it out. My test used the rhel6 and rhel7 main repos, publishing only rhel7. It takes an almost insignificant amount of time to publish the repo normally. With the diff above applied, it took my poor dev VM all night to finish the job. I didn't measure the times precisely, but the ratio appears to be "very bad", certainly well over the "unacceptably bad" line.

So there are a few different options here. The first, of course, is tolerate extremely slow errata publishes. I hate this option.

Another option is to use a different mechanism that can do the same filtering as in my diff, but waaaaaay faster. There are several[2] bits[3] of code[4] that do things like this on the import side that can be used as a reference for creating something a little more task-specific that gets the job done more quickly for a publish. I'm skeptical of this option, because this is a relational problem in a nonrelational system, and I suspect that there's little that we can do on a per-unit basis to make this faster. Nevertheless, it's probably worth trying a few different methods to see if one manages to get the job done in a timely manner.

The option that I think is most extreme is to do something similar to what I've done when faced with a similar problem, which is to use advanced mongodb features to make changes in one collection based on the contents of another[5]. I think this option might actually be worse than just waiting a few hours for the publish to happen, since it will involve writing more mongo-specific JS, which comes with a couple tons of baggage.

The remaining option might be the best one, which is one I haven't yet thought of. I welcome comments on this issue, and in the meantime will put this down for a little while so I can come back to it fresh and start trying the available options.

[0]: https://github.com/pulp/pulp_rpm/pull/625
[1]: https://bugzilla.redhat.com/show_bug.cgi?id=1171278
[2]: https://github.com/pulp/pulp_rpm/blob/43970f1f51302fba1db23483285acbcb6a885714/plugins/pulp_rpm/plugins/importers/yum/existing.py#L64-L89
[3]: https://github.com/pulp/pulp_rpm/blob/43970f1f51302fba1db23483285acbcb6a885714/plugins/pulp_rpm/plugins/importers/yum/associate.py#L90-L125
[4]: https://github.com/pulp/pulp/blob/ec59a2a794b1d59b9f88986952eea73e27be6803/server/pulp/plugins/conduits/mixins.py#L234-L255
[5]: https://github.com/pulp/pulp_rpm/blob/43970f1f51302fba1db23483285acbcb6a885714/plugins/pulp_rpm/plugins/importers/yum/purge.py#L293-L496

#### #5 - 03/07/2016 09:27 PM - mhrivnak

I like the track you're on. I think just modifying the search algorithm a touch will do it. FWIW a similar behavior already is happening in the profiler. See here: https://pulp.plan.io/issues/1366#note-5

Speaking of issue 1366, I think it's a dup of this one. I'll assign it to you and let you decide which to close.

Anyway, I bet this would be fast:

- do a pass through all the errata and collect the nevra
- do one search for all of those nevra in the repo
- do a second pass through the errata while actually writing the XML file, and filter out any packages that weren't found in the repo

#### #6 - 03/08/2016 07:48 PM - semyers

*- Is duplicate of Issue #1366: Advisory package list doesn't match packages in the repository added*

#### #7 - 03/10/2016 05:35 PM - semyers

*- Status changed from ASSIGNED to POST*

*- Version set to Master*

https://github.com/pulp/pulp_rpm/pull/825

#### #8 - 03/18/2016 07:10 PM - semyers

*- Status changed from POST to MODIFIED*

*- % Done changed from 0 to 100*

Applied in changeset 73d67dd4ac855b6311a5622f7eb211f2e65d0f35.

#### #9 - 03/23/2016 07:55 PM - semyers

*- Status changed from MODIFIED to 5*

#### #10 - 03/28/2016 10:58 PM - pthomas@redhat.com

*- Status changed from 5 to 6*

Verified
[root@ibm-x3550m3-10 ~]# rpm -qa |grep pulp-server
pulp-server-2.8.1-0.1.beta.el7.noarch
consumer only displaying packages in the repo

#### #11 - 04/05/2016 10:19 PM - semyers

*- Status changed from 6 to CLOSED - CURRENTRELEASE*

#### #13 - 04/15/2019 10:37 PM - bmbouter

*- Tags Pulp 2 added*