

Greg,

Here is an updated version of the sysfs tagged directories that improves a bit the situation over the previous one.

I've modified the patch 09 ("Enable tagging for net\_class directories in sysfs") to be a bit less intrusive in sysfs core. I removed the #ifdef'd parts you didn't like in fs/sysfs/mount.c, and replaced it by a generic routine `sysfs_ns_exit()` that is called, if needed, by the namespace when it exits. This routine goes through every sysfs super blocks and calls the callback passed by the namespace to clean its tag.

The patch is now splitted in two:

- \* 09/11: the generic routine,
- \* 10/11: the remaining network parts.

The generic part can be merge with patch 05 ("sysfs: Implement sysfs tagged directory support.") but I left it separate for now to ease reviews.

Serge's patch for user namespace is modified to use this new service too. No more #ifdef CONFIG\_NET or #ifdef CONFIG\_USER\_NS in fs/sysfs/mount.c now.

But, currently, a new namespace that wants to add its tag to sysfs dirs still need to modify fs/sysfs/mount.c in a few routines to manage the new tag member added in struct `sysfs_tag_info`: `sysfs_fill_super()`, `sysfs_test_super()`, `sysfs_kill_sb()` (see the last two patches). These changes are only the initialization and a bunch of comparisons.

If we really want to go further, to get rid of these, I've thought about:

- \* Extending `sysfs_tagged_dir_operations` with super blocks operations:
  - `fill_sb_tag`, `test_sb_tag`, `kill_sb_tag`
- \* Add routines in sysfs to allow registration/unregistration of these operations structs in a list:
  - `sysfs_register_tagged_dir_ops()`...
- \* Each subsystem concerned implements and registers its operations at boot.

\* In `sysfs_fill_super()`, `sysfs_test_super()` and `sysfs_kill_sb()`, add loops to go through all registered operations structs and calls the corresponding operations if it's present.

But... I thought it was a bit overkill for the few namespaces that will actually need sysfs tagged directories.

(Below you'll find the traditional introduction for sysfs tagged dirs and the updated changelog)

Thanks,  
Benjamin

--

Here is an updated version of Eric Biederman's patchset to implement tagged directories in sysfs ported on top of 2.6.26-rc2-mm1.

With the introduction of network namespaces, there can be duplicate network interface names on the same machine. Indeed, two network interfaces can have the same name if they reside in different network namespaces.

- \* Network interfaces names show up in sysfs.
- \* Today there is nothing in sysfs that is currently per namespace.
- \* Therefore we need to support multiple mounts of sysfs each showing a different network namespace.

We introduce tagged directories in sysfs for this purpose.

Of course the usefulness of this feature is not limited to network stuff: Serge Hallyn wrote a patch to fix a similar issue with user namespaces based on this patchset. His patch is included at the end of the patchset.

Tested with and without `SYSFS_DEPRECATED`. No regression found so far.

## Changelog

-----

- \* V5:
  - Make namespace tags a bit less intrusive in sysfs core:
  - New patch 09: Added a generic `sysfs_ns_exit` routine called by exiting namespaces. A callback is passed to this routine to execute the subsystem specific code.
  - Modified patches 09 and 10 (now 10 and 11) ("netns tagging" and "usersns tagging") to use this new routine instead of adding `#ifdef'd` code in `fs/sysfs/mount.c`.
  - Added missing `-ENOMEM` in `fs/sysfs/dir.c:prep_rename()` (Roel Kluin)
- \* V4:

- Ported to 2.6.26-rc2-mm1
  - Updated patch for user namespace by Serge Hallyn (patch 10).
- \* V3:
- Removed patch 10 ("avoid kobject name conflict with different namespaces"), a better one was provided by Eric.
  - Removed patch 11 ("sysfs: user namespaces: add ns to user\_struct"), Serge needs to rework some parts of it.
  - Change Acked-by: to Signed-off-by:, someone told me it is more appropriate (as I'm in the delivery path).

Here is the announcement Eric wrote back in December to introduce his patchset:

"

Now that we have network namespace support merged it is time to revisit the sysfs support so we can remove the dependency on !SYSFS.

[...]

The bulk of the patches are the changes to allow multiple sysfs superblocks.

Then comes the tagged directory sysfs support which uses information captured at mount time to decide which object with which tag will appear in a directory.

Then the support for renaming and deleting objects where the source may be ambiguous because of tagging.

Then finally the network namespace support so it is clear how all of this tied together.

"

Regards,  
Benjamin

--

---

Containers mailing list  
Containers@lists.linux-foundation.org  
<https://lists.linux-foundation.org/mailman/listinfo/containers>

---