Subject: userns: targeted capabilities v5
Posted by serge on Thu, 17 Feb 2011 15:02:24 GMT
View Forum Message <> Reply to Message

Here is a repost of my previous user namespace patch, ported onto
last night's git head.

It fixes several things I was doing wrong in the last (v4)
posting, in particular:

1. don't set uts_ns->user_ns to current's when !CLONE_NEWUTS
2. add a ipc_ns->user_ns which owns ipc_ns, and use that to
   decide CAP_IPC_OWNER
3. fix logic flaw caused by bad parantheses
4. allow do_prlimit to current
5. don't always give root full privs to init_user_ns

The expected course of development for user namespaces is laid out at https://wiki.ubuntu.com/UserNamespace. Bugs aside, this patchset is supposed to not at all affect systems which are not actively using user namespaces, and only restrict what tasks in child user namespace can do. They begin to limit privilege to a user namespace, so that root in a container cannot kill or ptrace tasks in the parent user namespace, and can only get world access rights to files. Since all files currently belong to the initila user namespace, that means that child user namespaces can only get world access rights to *all* files. While this temporarily makes user namespaces bad for system containers, it starts to get useful for some sandboxing.

I've run the 'runltplite.sh' with and without this patchset and found no difference. So all in all, this is the first version of this patchset for which I feel comfortable asking: please consider applying.

thanks,
-serge

____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containe rs