Subject: [RFC] [PATCH 0/3] user ns and vfs: Introduction
Posted by serue on Wed, 15 Nov 2006 17:40:19 GMT
View Forum Message <> Reply to Message

From: Serge E. Hallyn <serue@us.ibm.com>
Subject: [RFC] [PATCH 0/3] user ns and vfs: Introduction

Cedric has previously sent out a patchset
(http://lists.osdl.org/pipermail/containers/2006-August/000078.html)
impplementing the very basics of a user namespace. It ignores
filesystem access checks, so that uid 502 in one namespace could
access files belonging to uid 502 in another namespace, if the
containers were so set up.

This isn't necessarily bad, since proper container setup should
prevent problems. However there has been concern, so here is a
patchset which takes one course in addressing the concern.

This patchset adds assigns each vfsmount to the user namespace
of the process which did the mount. It introduces a userns-shared
mount flag mainly to allow a filesystem to be used by a container
while it is setting up. It could also be used along with read-only
bind mounts to share, for instance, /usr among mutiple containers.

This patchset replaces the previous one, which annotated the
superblock.

Is this direction in which we want to go? For instance, would we
want to allow the notion of a uidmap so that user 500 (hallyn)'s
files on the host system are owned by uid 0 in a container which
hallyn started? It's my impression that that could only be cleanly
done with either a stackable filesystem to give us fresh inodes
inside the container. Also, would a uidmap map uid's as stored on
disk to the mapped uids, or would we want to only support a uidmap
for the whole user namespace?

My own impression is that we are better off enfocing isolation than
trying to provide actual uid mapping, but please argue and discuss.

thanks,
-serge

____
Containers mailing list
Containers@lists.osdl.org
https://lists.osdl.org/mailman/listinfo/containers