Subject: Re: [PATCH 10/10] sysfs: user namespaces: add ns to user_struct
Posted by ebiederm on Wed, 30 Apr 2008 06:47:34 GMT
View Forum Message <> Reply to Message

"Serge E. Hallyn" <serue@us.ibm.com> writes:

>> > Index: linux-mm/include/linux/sched.h
>> > ===
>> > --- linux-mm.orig/include/linux/sched.h
>> > +++ linux-mm/include/linux/sched.h
>> > @@ -598,7 +598,7 @@ struct user_struct {
>> >
>> > /* Hash table maintenance information */
>> > struct hlist_node uidhash_node;
>> > - uid_t uid;
>> > + struct k_uid_t uid;
>>
>> If we are going to go this direction my inclination
>> is to include an array of a single element in user_struct.
>>
>> Maybe that makes sense. I just know we need to talk about
>> how a user maps into different user namespaces. As that
>
> My thought had been that a task belongs to several user_structs, but
> each user_struct belongs to just one user namespace. Maybe as you
> suggest that's not the right way to go.
>
> But are you ok with just sticking a user_namespace * in here for now,
> and making it clear that the user_struct-user_namespace relation is yet
> to be defined?
>
> If not that's fine, we just won't be able to clone(CLONE_NEWUSER)
> until we get the relationship straightened out.
>
>> is a real concept that really occurs in real filesystems
>> like nfsv4 and p9fs, and having infrastructure that can
>> deal with the concept (even if it doesn't support it yet) would be
>> useful.
>
> I'll have to look at 9p, bc right now I don't know what you're talking
> about. Then I'll move to the containers list to discuss what the
> user_struct should look like.

Ok. The concept present in nfsv4 and 9p is that a user is represented
by a username string instead by a numerical id. nfsv4 when it encounters
a username it doesn't have a cached mapping to a uid calls out to userspace to
get that mapping. 9p does something similar although I believe less general.

The key point here is that we have clear precedent of a mapping from one user namespace to another in real world code. In this case nfsv4 has one user namespace (string based) and the systems that mount it have a separate user namespace (uid based).

Once user namespaces are fleshed out I expect that same potential to exist. That each user namespace can have a different uid mapping for the same username string on nfsv4.

>From uid we current map to a user struct. At which point things get a little odd. I think we could swing either way. Either keeping kernel user namespaces completely disjoint or allowing them to be mapped to each other.

I certainly like the classic NFS case of mapping uid 0 to user nobody on a nonlocal filesystem (outside of the container in our case) so the don't accidentally do something that root only powers would otherwise allow.

In general I think managing mapping tables between user namespaces is a pain in the butt and something to be avoided if you have the option. I do see a small place for it though.

Eric