

---

Subject: Re: [RFC][PATCH 1/5] Virtualization/containers: startup  
Posted by [ebiederm](#) on Mon, 06 Feb 2006 09:19:37 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Kirill Korotaev <dev@sw.ru> writes:

>>> I think that a patch like this - particularly just the 1/5 part - makes total  
>>> sense, because regardless of any other details of virtualization, every  
>>> single scheme is going to need this.

>> I strongly disagree with this approach. I think Al Viro got it  
>> right when he created a separate namespace for filesystems.

> These patch set introduces separate namespaces as those in filesystems.  
> What exactly you don't like in this approach? Can you be more specific?

That you placed the namespaces in a separate structure from task\_struct.  
That part seems completely unnecessary, that and the addition of a  
global id in a completely new namespace that will be a pain to virtualize  
when it's time comes.

>> First this presumes an all or nothing interface. But that is not  
>> what people are doing. Different people want different subsets  
>> of the functionality. For the migration work I am doing having  
>> multiple meanings for the same uid isn't interesting.

> What do you mean by that? That you don't care about virtualization of UIDs? So  
> your migration doesn't care at all whether 2 systems have same uids? Do you keep  
> /etc/passwd in sync when do migration?

It is already in sync. When the only interesting target is inside of  
a cluster it is trivial to handle. In addition frequently there is  
only one job on a given machine at a time so there is no other  
interesting user of UIDs.

> Only full virtualization allows to migrate applications without bugs and  
> different effects.

The requirement is that for every global identifier you have a  
namespace where you will not have a conflict in definition when you  
restore that identifier. (Either that or you have to update all  
references to that identifier which is usually too hideous to take  
seriously)

Virtualization has nothing to do with it. It is all about  
non-conflicting names. The kernel already has everything virtualized.

>> Secondly by implementing this in one big chunk there is no  
>> migration path when you want to isolate an additional part of the

>> kernel interface.  
>> So I really think an approach that allows for incremental progress  
>> that allows for different subsets of this functionality to  
>> be used is a better approach. In clone we already have  
>> a perfectly serviceable interface for that and I have  
>> seen no one refute that. I'm not sure I have seen anyone  
>> get it though.  
> Just introduce config option for each virtualization functionality. That's it.

If you have different pointers for each one that is sane. I meant in the api. I did not see patches 4/5 and 5/5. But I presumed since you allocated a great global structure everything was turned on off with that.

>> My apologies for the late reply I didn't see this thread until  
>> just a couple of minutes ago. linux-kernel can be hard to  
>> follow when you aren't cc'd.  
>> Patches hopefully sometime in the next 24hours. So hopefully  
>> conversation can be carried forward in a productive manner.  
> Ok. I will remake them either :)

My not seeing your user space api is one of the problems with incomplete patches which seem to plague this conversation.

Eric

---