Subject: Re: [RFC][PATCH 3/4]: Enable multiple mounts of /dev/pts
Posted by Pavel Emelianov on Wed, 06 Feb 2008 16:56:37 GMT
View Forum Message <> Reply to Message

[snip]

>> Mmm. I wanted to send one small objection to Cedric's patches with mqns,
>> but the thread was abandoned by the time I decided to do-it-right-now.
>>
>> So I can put it here: forcing the CLONE_NEWNS is not very good, since
>> this makes impossible to push a bind mount inside a new namespace, which
>> may operate in some chroot environment. But this ability is heavily
>
> Which direction do you want to go? I'm wondering whether mounts
> propagation can address it.

Hardly. AFAIS there's no way to let the chroot-ed tasks see parts of
vfs tree, that left behind them after chroot, unless they are in the
same mntns as you, and you bind mount this parts to their tree. No?

> Though really, I think you're right - we shouldn't break the kernel
> doing CLONE_NEWMQ or CLONE_NEWPTS without CLONE_NEWNS, so we shouldn't
> force the combination.
>
>> exploited in OpenVZ, so if we can somehow avoid forcing the NEWNS flag
>> that would be very very good :) See my next comment about this issue.
>>
>>> Pavel, not long ago you said you were starting to look at tty and pty
>>> stuff - did you have any different ideas on devpts virtualization, or
>>> are you ok with this minus your comments thus far?
>> I have a similar idea of how to implement this, but I didn't thought
>> about the details. As far as this issue is concerned, I see no reasons
>> why we need a kern_mount-ed devtpsfs instance. If we don't make such,
>> we may safely hold the ptsns from the superblock and be happy. The
>> same seems applicable to the mqns, no?
>
> But the current->nsproxy->devpts->mnt is used in several functions in
> patch 3.

Indeed. I overlooked this. Then we're in a deep ... problem here.

Breaking this circle was not that easy with pid namespaces, so
I put the strut in proc_flush_task - when the last task from the
namespace exits the kern-mount-ed vfsmnt is dropped, but we can't
do the same stuff with devpts.

I do not remember now what the problem was and it's already quite
late in Moscow, so if you don't mind I'll revisit the issue tomorrow.

Off-topic: does any of you know whether Andrew is willing to accept
new features in the nearest future? The problem is that I have a
device visibility controller fixed and pending to send, but I can't
guess a good time for it :)

>> The reason I have the kern_mount-ed instance of proc for pid namespaces
>> is that I need a vfsmount to flush task entries from, but allowing
>> it to be NULL (i.e. no kern_mount, but optional user mounts) means
>> handing all the possible races, which is too heavy. But do we actually
>> need the vfsmount for devpts and mqns if no user-space mounts exist?
>>
>> Besides, I planned to include legacy ptys virtualization and console
>> virtualizatin in this namespace, but it seems, that it is not present
>> in this particular one.
>
> I had been thinking the consoles would have their own ns, since there's
> really nothing linking them, but there really is no good reason why
> userspace should ever want them separate. So I'm fine with combining
> them.

OK.