Subject: Re: [RFC] ns containers (v2): namespace entering
Posted by Paul Menage on Thu, 22 Feb 2007 16:53:10 GMT
View Forum Message <> Reply to Message

On 2/22/07, Eric W. Biederman <ebiederm@xmission.com> wrote:
>
> The really important use of the ptrace case is that it works using
> existing mechanisms without leaks. So it is very useful yardstick.
>
> The other important yardstick is arranging it so that when you login
> to a machine all of the user code runs in your target environment.
> How you get there is irrelevant.
>
> One of the cases I have been worrying about in looking at the
> semantics of enter is what do you do with the parent pid. Supporting
> ptrace from outside the pid namespace of a process inside a pid
> namespace requires supporting a parent process outside of the pid
> namespace for processes other than init.
>

When I implemented a virtual server solution at my previous job, we solved the problem of leaking capabilities into the virtualized environment by allowing a process to enter the virtual server in a "privileged" mode, in which it didn't appear in the virtualized /proc, and hence none of its resources were accessible to processes in the VS. Children of a privileged process were by default regular members of the virtual server.

With the nsproxy model, maybe you could implement that by having two pid namespaces. When you create the VS, you create an entire set of new namespaces; then before forking init you create a new pid_ns. So by entering the outer nsproxy you'd get access to the same environment that the VS sees, but your process wouldn't be visible to processes in the VS. If you wanted to create a regular process, you'd just enter the inner pid_ns too. I think that doing the appropriate process "sanitization" to avoid leaking capabilities would be easier from the "twilight zone" intermediate nsproxy than from the machine top-level.

Paul

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