

---

Subject: Re: [RFC PATCH 3/5] use next syscall data to predefine process ids  
Posted by [ebiederm](#) on Thu, 10 Jul 2008 09:36:19 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Nadia Derby <[Nadia.Derbey@bull.net](mailto:Nadia.Derbey@bull.net)> writes:

> If I correctly understood what you're saying, it means set min = max =  
> target\_pid using /proc/sys, i.e. for the whole system: don't you think this  
> might be dangerous: allocating pids will fail for any other running process  
> during the entire period of time where /proc/sys will be set like that.  
> I really think this is a feature that should be confined to a process.

Well for a pid namespace, so that is more confined.

Grr. We still need to move /proc/sys into /proc/<pid>/sys so it is clear that sysctls are per namespace.

You are right that doing it that way has downsides. In particular it is hard to parallelize the restoration of a pid namespace.

However the interface does exist, and it didn't look like you were reusing that code in your allocator.

It is my firm suspicion that restoring a process one syscall at a time is too fine a granularity. Certainly for the VM of a process it is.

So here is my suggestion for now. Take whatever approach you are doing and make it work for you. Go as far as you can go and see what the pitfalls are. Then on the 22nd we can all get in a room and discuss things, and if we are lucky agree on a path forward.

Eric

---

Containers mailing list  
[Containers@lists.linux-foundation.org](mailto:Containers@lists.linux-foundation.org)  
<https://lists.linux-foundation.org/mailman/listinfo/containers>

---