

---

Subject: Re: [RFC][PATCH] Add child reaper to struct pspace

Posted by [dev](#) on Sat, 16 Sep 2006 11:55:04 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Eric W. Biederman wrote:

> Kirill Korotaev <[dev@sw.ru](mailto:dev@sw.ru)> writes:

>

>

>

>>I guess there will be a need of list of tasks... not of pids only...

>>many of loops like `do_each_thread()/while_each_thread()` has nothing to do with

>>pids

>>and should be narrowed down to loop through the container.

>>

>>Does this logic belong to `pid_ns`? if yes, then it definitely should be called

>>`task_ns`.

>

>

> Just skimming through I see one or two instances. Where the existing  
> loop uses `do_each_thread()/while_each_thread()` that we need to change.

>

> `kernel/capabilities.c cap_set_all()` is an example.

>

> However what we are trying to achieve there is to iterate through

> the same list that `kill(-1, )` uses. So we need to replace

> `do_each_thread()/while_each_thread()` with something that will

> iterate through everything in the pid namespace.

>

> Most instances of `do_each_thread()/while_each_thread()` the kernel

> really does need a global view, and need to be left unchanged.

>

> Basically the current kernel is short the concept of a process

> group of all processes, and uses the concept of a list of all processes

> instead.

>

> Since the two concepts of a list of all tasks, and a list of all processes

> I can see diverge when we have multiple pid namespaces we need to add

> an additional concept, in the kernel.

>

> Do you know an example in that we need to change to implement a pid

> namespace that goes beyond iterating through the list of processes

> that `kill(-1,)` uses?

from OVZ patches:

`do_each_thread_ve()`

`elf_core_dump()` (need pid namespace list?)

`zap_threads` (need pid namespace list?)

chroot\_fs\_refs  
cap\_set\_all (need pid namespace list?)  
cpt functions (need to freeze VE processes, pid namespace list?)  
sys\_setpriority (needs task list for user namespace!)  
sys\_getpriority (the same)  
sys\_ioprio\_set (the same)  
sys\_ioprio\_get (the same)  
selinux\_setprocattr (should be changed with the check for thread\_group\_empty()???)

for\_each\_process\_ve()  
  asids\_proc\_info (need pid namespace list? in host should print all?)  
  kill\_something\_info (I suppose you changed it already?)

some of these are optimizations which are natural for containers and are good for scalability (as zap\_threads, elf\_core\_dump etc.).

Thanks,  
Kirill  
P.S. Sorry for not always replying quickly...

---