
Subject: Re: containers development plans

Posted by [Cedric Le Goater](#) on Thu, 05 Jul 2007 15:53:37 GMT

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

Hello,

some more comments on what we talked about at OLS and what we are working on.

Serge E. Hallyn wrote:

- > We are trying to create a roadmap for the next year of
- > 'container' development, to be reported to the upcoming kernel
- > summit. Containers here is a bit of an ambiguous term, so we are
- > taking it to mean all of:
- >
- > 1. namespaces
- > 2. process containers
- > 3. checkpoint/restart
- >
- > Naturally we can't actually predict what will and won't be worked on,
- > let alone what will be going upstream. But the following is a list
- > of features which it seems reasonable to think might be worked on
- > next year:
- >
- > 1. completion of ongoing namespace

the ipc namespace would need a "set identifier" feature if we were to use it for C/R. this is not available right now. a patchset was sent introducing a new IPC_SETID but it didn't get much attention.

- > pid namespace

At OLS, we agreed that suka's hierarchical pidns patchset should be fine if we can make sure perms are OK when the namespace is not used. right ?

I get < 1% today, so it should be okay :)

There are still some issues around /proc that we are working on. Hopefully, we should be able to merge most of the helpers patch real soon.

we need a clone_with_pid() kind of syscall for C/R. I had planned to work on a :

```
clone64(struct clone64_arg_struct *arg)
```

to extend the clone flags which will soon overflow. we could

easily add a pid attribute to implement the clone_with_pid()

the kthread cleanup is not completed yet. some patch are pending but i would say that the most important ones are around NFS and i'm not sure anyone worked on these.

af unix credentials still hold some pid_t's. they need a clean up.

> net namespace

see previous email

> ro bind mounts

work in progress. dave ?

what about mounting /proc and /sys multiple times ?

> 2. continuation with new namespaces
> devpts, console, and ttydrivers
> user

merged experimental. we still need to work on the (user,usersns) checks. however, openvz and linux-vserver should already be able to use it.

> time
> namespace management tools
> namespace entering

there are a few patchsets on the topic :

- * bind_ns() syscall
- * container subsystem identifying a nsproxy object

but they didn't get much review :(

> 3. any additional work needed for virtual servers?
> i.e. in-kernel keyring usage for cross-username namespace permissions, etc
> 4. task containers functionality
> base features
> specific containers
> poll to see who has plans
> 5. checkpoint/restart

we really need to leverage the freezer and suspend to disk for that. there are some talks about it right now but it seems a bit early to have clear directions yet.

generalizing the refrigerator to all arch seems a good idea to freeze a container. then how do we initiate checkpoint ? syscall ? signal ? etc.

These topics were addressed at the BOF and people are now aware of different solutions. we hope that the email storm on what directions to take for mainline will start soon.

C.

- > memory c/r
- > (there are a few designs and prototypes)
- > (though this may be ironed out by then)
- > per-container swapfile?
- > overall checkpoint strategy
- > in-kernel vs userspace-driven

hybrid ?

- > overall restart strategy
- >
- > What more needs to be added to this list?
- >
- > A list of the people we are currently aware of who are showing interest
- > in these features follows. What I'd like to know is, from this list, do
- > some people know what general or specific areas they plan to or want to
- > work on over the next year?
- >
- > Stakeholders:
- > Eric Biederman
- > google
- > ibm
- > kerlabs
- > openvz
- > osdl (Masahiko Takahashi?)
- > vserver
- >
- > Who is missing from the list?
- >
- > thanks,
- > -serge
- >
- > _____
- > Containers mailing list
- > Containers@lists.linux-foundation.org
- > <https://lists.linux-foundation.org/mailman/listinfo/containers>
- >

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>
