Subject: [PATCHSET] 2.6.20-rc4-mm1-lxc2
Posted by Cedric Le Goater on Tue, 16 Jan 2007 17:41:01 GMT
View Forum Message <> Reply to Message

All,

We've been gathering and porting patches related to namespaces in
a lxc patchset for a while now. Mostly working on the network
namespace which will require some extra work to be usable.

 * It's available here :

 http://www.sr71.net/patches/2.6.20/2.6.20-rc4-mm1-lxc2/

* Caveats :

  namespace syscalls are still under construction.

  network namespace is broken :

  . the nsproxy backpointer in net_ns is flaky.
  . the push_net_ns() and pop_net_ns() can be called under
    irq and are using current. this seems inappropriate.
  . there is a race on ->nsproxy between push_net_ns() and
    exit_task_namespaces()
  . does not compile with CONFIG_NET_NS=n

  pid namespace is still under construction.

  ro bind mounts should be pushed soon

thanks,

C.

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

Subject: Re: [PATCHSET] 2.6.20-rc4-mm1-lxc2
Posted by Daniel Lezcano on Tue, 16 Jan 2007 23:48:27 GMT
View Forum Message <> Reply to Message

Cedric Le Goater wrote:
> All,
>
> We've been gathering and porting patches related to namespaces in

> a lxc patchset for a while now. Mostly working on the network
> namespace which will require some extra work to be usable.
>
> * It's available here :
>
> http://www.sr71.net/patches/2.6.20/2.6.20-rc4-mm1-lxc2/
>
> * Caveats :
>
> namespace syscalls are still under construction.
>
> network namespace is broken :
>
> . the nsproxy backpointer in net_ns is flaky.
> . the push_net_ns() and pop_net_ns() can be called under
>   irq and are using current. this seems inappropriate.
> . there is a race on ->nsproxy between push_net_ns() and
>   exit_task_namespaces()

Hi Dmitry,

we are experiencing NULL address access when using the nsproxy in
push_net_ns function without any unshare.

It appears the exit_task_namespace function sets current->nsproxy to
NULL and we are interrupted by an incoming packet. The netif_receive_skb
does push_net_ns(dev->net_ns). The push_net_ns function retrieves the
current->nsproxy to use it. But it was previously set to NULL by the
exit_task_namespace function.

The bug can be reproduced with the following command launched from
another host.

while $(true); do ssh myaddress ls > /dev/null && echo -n .; done

After a time (between 1 second - 3 minutes), the kernel panics.

I think this will be very hard to fix and perhaps we should redesign
some part. Instead of using nsproxy swapping, perhaps we should pass
net_ns as parameter to functions, but that will breaks a lot of API.

What is your feeling on that ?

Regards.

  -- Daniel.

_____
Containers mailing list

Subject: Re: [PATCHSET] 2.6.20-rc4-mm1-lxc2
Posted by ebiederm on Wed, 17 Jan 2007 01:46:35 GMT
View Forum Message <> Reply to Message

Daniel Lezcano <dlezcano@fr.ibm.com> writes:

>
> Hi Dmitry,
>
> we are experiencing NULL address access when using the nsproxy in
> push_net_ns function without any unshare.
>
> It appears the exit_task_namespace function sets current->nsproxy to
> NULL and we are interrupted by an incoming packet. The netif_receive_skb
> does push_net_ns(dev->net_ns). The push_net_ns function retrieves the
> current->nsproxy to use it. But it was previously set to NULL by the
> exit_task_namespace function.
>
> The bug can be reproduced with the following command launched from
> another host.
>
> while $(true); do ssh myaddress ls > /dev/null && echo -n .; done
>
> After a time (between 1 second - 3 minutes), the kernel panics.
>
> I think this will be very hard to fix and perhaps we should redesign
> some part. Instead of using nsproxy swapping, perhaps we should pass
> net_ns as parameter to functions, but that will breaks a lot of API.
>
> What is your feeling on that ?

After looking at several things primarily ramifications of file descriptor
passing I have concluded that a magic global variable in the task struct
is almost certainly the wrong thing to do. And the more I look at it
the task is usually the wrong location to look to see what network namespace
you are in.

To that effect I have been preparing a patchset for discussion targeting
the end of this week to have it ready, in an easily reviewable format.

Eric

_____

Containers mailing list
Containers@lists.osdl.org

Subject: Re: [PATCHSET] 2.6.20-rc4-mm1-lxc2
Posted by Mishin Dmitry on Wed, 17 Jan 2007 10:57:10 GMT
View Forum Message <> Reply to Message

On Wednesday 17 January 2007 02:48, Daniel Lezcano wrote:
> Cedric Le Goater wrote:
> > All,
> >
> > We've been gathering and porting patches related to namespaces in
> > a lxc patchset for a while now. Mostly working on the network
> > namespace which will require some extra work to be usable.
> >
> > * It's available here :
> >
> > http://www.sr71.net/patches/2.6.20/2.6.20-rc4-mm1-lxc2/
> >
> > * Caveats :
> >
> > namespace syscalls are still under construction.
> >
> > network namespace is broken :
> >
> > . the nsproxy backpointer in net_ns is flaky.
> > . the push_net_ns() and pop_net_ns() can be called under
> > irq and are using current. this seems inappropriate.
> > . there is a race on ->nsproxy between push_net_ns() and
> > exit_task_namespaces()
>
> Hi Dmitry,
>
> we are experiencing NULL address access when using the nsproxy in
> push_net_ns function without any unshare.
>
> It appears the exit_task_namespace function sets current->nsproxy to
> NULL and we are interrupted by an incoming packet. The netif_receive_skb
> does push_net_ns(dev->net_ns). The push_net_ns function retrieves the
> current->nsproxy to use it. But it was previously set to NULL by the
> exit_task_namespace function.
>
> The bug can be reproduced with the following command launched from
> another host.
>
> while $(true); do ssh myaddress ls > /dev/null && echo -n .; done
>
> After a time (between 1 second - 3 minutes), the kernel panics.

>
> I think this will be very hard to fix and perhaps we should redesign
> some part. Instead of using nsproxy swapping, perhaps we should pass
> net_ns as parameter to functions, but that will breaks a lot of API.
I've redesigned this already to use per-CPU global variable, as Eric
suggests. Updated l2 networking patchset will be sent later today or tommorow.
Sorry for the latency, there were very long holidays here :)

--
Thanks,
Dmitry.

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

Subject: Re: [PATCHSET] 2.6.20-rc4-mm1-lxc2
Posted by Daniel Lezcano on Wed, 17 Jan 2007 11:18:45 GMT
View Forum Message <> Reply to Message

Dmitry Mishin wrote:

[cut]

>> I think this will be very hard to fix and perhaps we should redesign
>> some part. Instead of using nsproxy swapping, perhaps we should pass
>> net_ns as parameter to functions, but that will breaks a lot of API.

> I've redesigned this already to use per-CPU global variable, as Eric
> suggests. Updated l2 networking patchset will be sent later today or tommorow.
> Sorry for the latency, there were very long holidays here :)

The longer they are, the best it is ;)

BTW, did you fix the CONFIG_NET_NS=n compilation ?

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