Subject: Re: [PATCH 1/12] L2 network namespace: current network namespace operations
Posted by Herbert Poetzl on Sat, 09 Dec 2006 04:24:10 GMT
View Forum Message <> Reply to Message

On Fri, Dec 08, 2006 at 01:03:29PM -0700, Eric W. Biederman wrote:
> Dmitry Mishin <dim@openvz.org> writes:
>
> > On Thursday 07 December 2006 12:37, Cedric Le Goater wrote:
> >> Dmitry Mishin wrote:
> >> > Added functions and macros required to operate with current network
> > namespaces.
> >> > They are required in order to switch network namespace for incoming packets
> > and
> >> > to not extend current network interface by addtiotional network namespace
> > argue.
> >> >
> >> > Signed-off-by: Dmitry Mishin <dim@openvz.org>
> >> >
> >> > ---
> >> > include/linux/net_namespace.h | 41 +++++++++++++++++++++++++++++++++++++++--
> >> > kernel/nsproxy.c | 1 +
> >> > net/core/net_namespace.c | 2 ++
> >> > 3 files changed, 42 insertions(+), 2 deletions(-)
> >> >
> >> > --- linux-2.6.19-rc6-mm2.orig/include/linux/net_namespace.h
> >> > +++ linux-2.6.19-rc6-mm2/include/linux/net_namespace.h
> >> > @@ -7,6 +7,7 @@
> >> >
> >> > struct net_namespace {
> >> >   struct kref kref;
> >> > + struct nsproxy *ns;
> >> > };
> >> >
> >>
> >> why do you need that back pointer ?
> >>
> >> (The answer must be in the following patches but I'm being lazy and
> >>   asking the author :)
> > Because for the incoming packets, I need to switch networking
> > namespace per-task and not per-nsproxy. If I switch it just in
> > current->nsproxy, it means that all tasks shring this nsproxy will
> > switch network context.
> >
> > It is one of the reasons, why we need per-task exec_context pointer.
>
> Ugh.
>

> If necessary push_net_ns and pop_net_ns should be implemented like
> get_fs and set_fs. Using a sane variable in thread_info, or per cpu.
>
> I'm not at all certain I'm comfortable doing this in interrupt
> context.

it should not be necessary to do that, and IMHO
changing the namespace temporarily is not such
a good idea, as that might cause all kinds of
ugly races, when other parts of the OS (from
other CPUs) access process relevant information
(utilizing the namespaces)

> Assuming we are doing it then we should do it for every path both
> socket and network device and do the lookup once and the cache it
> globally in the current execution context.
>
> We should not change current->nsproxy. I don't think for packet
> processing we need to change every namespace do we? The uid namespace
> and the like should be irrelevant correct?

hmm, wouldn't it be better to pass the relevant
information (network context) within the network
stack where needed, instead of changing the
network assignment of 'current' for processing
network packets?

I remeber from a prototype Linux-VServer implementation
that this wasn't that complicated to do ...

best,
Herbert

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