Subject: Re: net namespace plans for 2.6.25 (was Re: Pid namespaces problems)
Posted by den on Thu, 08 Nov 2007 14:04:41 GMT
View Forum Message <> Reply to Message

Daniel Lezcano wrote:
> Denis V. Lunev wrote:
>> Daniel Lezcano wrote:
>>> Denis V. Lunev wrote:
>>>> Daniel Lezcano wrote:
>>>>
>>>>> * the first one is the locking of the network namespace list by
>>>>> rtnl_lock, so from the timer callback we can not browse the network
>>>>> namespace list to check the age of the routes. It is a problem I would
>>>>> like to talk with Denis if he has time
>>>> From my point of view, the situation is clear. The timer should be
>>>> per/namespace. The situation is completely different as one in IPv4.
>>> We thought to make a timer per namespace for ipv6, but we are a little
>>> afraid for the performances when there will be a lot of containers.
>>> Anyway, we can do a timer per namespace and optimize that later. I will
>>> cook a new patch to take into account that for the next week.
>>
>> IMHO not a problem. tcp_write_timer is per/socket timer. If this works
>> efficiently, per/namespace one will work also.
>
> That's right, this is a good argument. By the way, the amount of work to
> be done in the tcp_write_timer is perhaps smaller than the one done in
> the ipv6 routing age check, no ? Anyway, I'm not against a timer per
> namespace in this case, I already did a try before rolling back to a
> for_each_net in the gc timer, that changes a little the API, but nothing
> we can handle easily.
>
>
I think you are wrong. The amount of work to "purge" all namespaces is a
constant in the IPv6 case, where we'll have per/namespace cache. So, for
a multiple timers model only multiple timer overhead counts and this
overhead is small, as timer list is efficient.

This argument does not for for IPv4 case, where there is a one big cache
for all namespaces.

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