Subject: Re: [PATCH][NETNS] Make ifindex generation per-namespace
Posted by Pavel Emelianov on Wed, 10 Oct 2007 08:55:32 GMT
View Forum Message <> Reply to Message

Eric W. Biederman wrote:
> Pavel Emelyanov <xemul@openvz.org> writes:
>
>> Currently indexes for netdevices come sequentially one by
>> one, and the same stays true even for devices that are
>> created for namespaces.
>>
>> Side effects of this are:
>> * lo device has not 1 index in a namespace. This may break
>>   some userspace that relies on it (and AFAIR something
>>   really broke in OpenVZ VEs without this);
>
> As it happens lo hasn't been registered first for some time
> so it hasn't had ifindex of 1 in the normal kernel.
>
>> * after some time namespaces will have devices with indexes
>>   like 1000000 os similar. This might be confusing for a
>>   human (tools will not mind).
>
> Only if we wind up creating that many devices.

Nope. Create and destroy new net ns for 10000 times and you'll get it.

>> So move the (currently "global" and static) ifindex variable
>> on the struct net, making the indexes allocation look more
>> like on a standalone machine.
>>
>> Moreover - when we have indexes intersect between namespaces,
>> we may catch more BUGs in the future related to "wrong device
>> was found for a given index".
>
> Not yet.
>
> I know there are several data structures internal to the kernel that
> are indexed by ifindex, and not struct net_device *. There is the
> iflink field in struct net_device. We need a way to refer to network
> devices in other namespaces in rtnetlink in an unambiguous way. I
> don't see any real problems with a global ifindex assignment until
> we start migrating applications.
>
> So please hold off on this until the kernel has been audited and
> we have removed all of the uses of ifindex that assume ifindex is
> global, that we can find.

Ok.

> Right now a namespace local ifindex seems to be just asking for
> trouble.

You said the same about caching the global pid on the task_struct,
but looks like you were wrong ;) Just kidding.

> Eric
>
>