Subject: Re: [PATCH 1/5] net: Modify all rtnetlink methods to only work in the initial namespace
Posted by den on Wed, 10 Oct 2007 14:29:05 GMT
View Forum Message <> Reply to Message

Daniel Lezcano wrote:
> struct net *net = in?in->nd_net:out->nd_net;
>
>> So, we are bound to the following options:
>> - perform additional non-uniform hacks around to place 'struct net' into
>>   other and other structures like xt_target
>> - add 7th parameter here and over
>> - introduce an skb_net field in the 'struct sk_buff' making all code
>>   uniform, at least when we have an skb
>>
>> I think that this is not the last place with such a parameter list and
>> we should make a decision at this point when the code in not mainline
>> yet.
>>
>> As far as I understand, netfilters are not touched by the Eric and we
>> can face some non-trivial problems there.
>
> In Eric's git tree:
> http://git.kernel.org/?p=linux/kernel/git/ebiederm/linux-2.6-netns.git
>
> There are some modifications concerning
> net/ipv4/netfiler/iptable_filter.c and at the ipt_hook function, there is:
>
> struct net *net = (in?in:out)->nd_net;
>
>> So, if my point about uniformity is valid, this patchset looks wrong and
>> should be re-worked :(
>
> As Eric said, we want to build the network namespace step by step,
> taking care of not breaking the init network namespace.
>
> If you want to make iptables per namespace or catch problems before the
> code goes to Dave's tree, IMHO it will be more convenient to post to
> containers@ the patches against netns49, where the modifications will be
> in a network namespace big picture.
>

my point is somewhat another. Yes, this is enough for that place. If so,
I must scatter these checks all around in the netfilters code. Brr.

In forward chain the situation is different for Layer3 switching. Let's
assume that we have an OpenVZ scheme, where the packet flows from socket
to device and after that from device to device via forwarding path. You

can't call skb_orphan on namespace switching as this breaks UDP flow regulation. Virtual network device is fast while real Ethernet is slow, packets will be dropped on queue in real device. So, the situation with packet on send path with a socket from other namespace is possible :(

I just pray for uniformity to concentrate on the code rather than on guesses on which path we are :(

Regards,
Den