Subject: Re: [PATCH 0/5] Make nicer CONFIG_NET_NS=n case code
Posted by Eric Dumazet on Wed, 31 Oct 2007 22:40:59 GMT
View Forum Message <> Reply to Message

> Eric Dumazet <dada1@cosmosbay.com> writes:
>
>
>> Definitly wanted here. Thank you.
>> One more refcounting on each socket creation/deletion was expensive.
>
> Really? Have you actually measured that? If the overhead is
> measurable and expensive we may want to look at per cpu counters or
> something like that. So far I don't have any numbers that say any
> of the network namespace work inherently has any overhead.

It seems that on some old opterons (two 246 for example),
"if (atomic_dec_and_test(&net->count))" is rather expensive yes :(

I am not sure per cpu counters help : I tried this and got no speedup. (This
was on net_device refcnt at that time)

(on this machines, the access through fs/gs selector seems expensive too)

Maybe a lazy mode could be done, ie only do a atomic_dec(), as done in dev_put() ?

Also, "count" sits in a cache line that contains mostly read and shared
fields, you might want to put it in a separate cache line in SMP, to avoid
cache line ping-pongs.

>
>> Maybe we can add a macro to get nd_net from a "struct net_device"
>> so that every instance of
>>
>> if (dev->nd_net != &init_net)
>>   goto drop;
>>
>> can also be optimized away if !CONFIG_NET_NS
>
> Well that extra check should be removed once we finish converting
> those code paths. So I'm not too worried.

OK. Since the conditional test can be predicted by cpu, it certainly doesnt
matter.

>
> If this becomes a big issue I can dig up my old code that

> replaced struct net * with a net_t typedef and used functions
> for all of the comparisons and allowed everything to be compiled
> away.

>
> Trouble was it was sufficiently different that it was just enough
> different that people could not immediately understand the code.
>
